

# Secure authentication watermarking for localization against the Holliman–Memon attack

Niladri B. Puhan · Anthony T. S. Ho

Published online: 16 November 2006  
© Springer-Verlag 2006

**Abstract** Authentication watermarking schemes using block-wise watermarks for tamper localization are vulnerable to the Holliman–Memon attack. In this paper, we propose a novel method based on the Wong’s localization scheme (Proceedings of the IS&T PIC, Portland) to resist this attack. A unique image index scheme is used for computing the authentication signature that is embedded in the least significant bit-plane of the block. The informed detector estimates the correct image index by using the side information about the watermarked image. The image index estimation from the fake image can definitely be an alternative to keeping a directory of image indices. So it is not necessary to manage the database of image indices for the verification purpose. The authenticity measure is defined to quantify the attack severity by taking the connectivity among possible authentic blocks into consideration. There are more blocks verified as authentic when this measure is high for a fake image constructed using this attack. As such, the blocks for a fake image can be chosen from a reduced number of database images. The blocks from any such image are to be connected with each other to maximize the authenticity measure. Thus, the attacker’s task to generate a fake image of reasonable perceptual quality

becomes increasingly difficult. With the proposed method there is no loss or ambiguity in localization after the Holliman–Memon attack and content tampering in an image. The localization accuracy in the proposed method is demonstrated by the simulation results and is equal to the chosen block size, similar to the Wong’s scheme.

**Keywords** Digital watermarking · Fragile authentication · Security · Localization · Holliman–Memon attack · Authenticity score

## 1 Introduction

Digital watermarking is the art of protecting the multimedia data by inserting the proprietary mark which may be easily retrieved by the owner of the data to verify about its ownership or authenticity. A watermark is desirable to be imperceptible to the human eye, secure against malicious attacks and fragile or robust depending on the type of application it addresses. A variety of digital watermarking methods have been developed for such purposes [1,2]. For certain applications, watermarks for checking the authenticity of the multimedia data should be fragile because any corruption to watermarked data easily destroy the watermark and so the detection algorithm will be able to verify the integrity of the data being tested.

Integrity and authenticity of digital media can be guaranteed through the use of cryptographic techniques to design the fragile watermark. In authentication applications, various algorithms such as DSA, RSA are used to generate the digital signature while algorithms such as SHA, MD5 are used to generate the hashed message

N. B. Puhan (✉)  
Center for Information Security,  
School of Electrical and Electronic Engineering,  
Nanyang Technological University,  
Singapore 639798, Singapore  
e-mail: niladri@pmail.ntu.edu.sg

A. T. S. Ho  
Department of Computing,  
School of Electronics and Physical Sciences,  
University of Surrey, Guildford, Surrey, UK  
e-mail: a.ho@surrey.ac.uk

authentication code (HMAC) [3]. In fragile authentication watermarking, the advantage of having the digital signature or HMAC hidden inside the digital data rather than appended to it is obvious. Lossless format conversion of the watermarked data does not render it inauthentic though the representation of the data is changed. Another advantage is that if the authentication information is localized, it is then possible to achieve the capability to localize the modifications after tampering by a hostile attacker. Localization is useful because knowledge of when and where the data has been altered can be used to infer the motive for tampering and identifying the culprits responsible.

In literature, there are two closely related approaches to tamper localization. The first, block-wise authentication divides an image into non-overlapping blocks and embeds an authentication watermark into each block independently. One of the first block-wise tamper localization schemes was proposed by Wong [4]. In this scheme, an image is divided into non-overlapping blocks and the watermarking is performed for each block independently. The seven most significant bits (MSBs) of all pixels in a block are hashed using a secure key-dependent hash. The hash is then XORed with a chosen binary logo and inserted into the LSBs of the same block. The watermark verification process starts in the reverse order by calculating the key-dependent hash of the seven MSBs in each block and XOR operation is performed with the LSBs. The tampered blocks can be found by comparing the output with the used logo. A public key version of this localization method is also suggested in [5]. The second, sample-wise authentication is an extreme case of block-wise authentication in which each block is reduced to the size of each sample [6].

In a fragile authentication scheme, the goal of the attacker is to modify the watermarked image(s) such that the resulting fake image can be verified as authentic. Several types of attacks are applicable to an authentication scheme depending on the application scenario. An interesting discussion is given about various attacks such as stego-image attack, multiple stego-image attack, verification device attack, cover-image attack and chosen cover-image attack in [7]. It has been analyzed that block-wise localization method is more secure than the second approach to resist hostile attacks [7]. Though the localization accuracy and cryptographic security offered by a localization method is high, its block-wise independence was used by Holliman and Memon [8] to design a counterfeiting attack. If a set of images are watermarked with the same key, it is possible to modify an arbitrary image to be authentic using this attack. The attacker divides the image into non-overlapping blocks and for each block performs a search in the set of

authentic blocks. The original block is replaced with the most similar block to maintain perceptual quality of the forged image. Thus the attacker creates the forged image by a collage of authentic blocks and the forged image is authenticated using block-wise independent watermarks. This particular counterfeiting attack is known as the Holliman–Memon attack or collage attack or vector quantization attack. Another weakness of Wong's scheme is that if the blocks in an image are swapped with each other, then the detector will still verify the modified image as authentic. A number of countermeasures have been proposed in the literature to resist this attack [7, 9–11]. However, most of these methods can resist this attack at the cost of localization accuracy. In this paper we suggest a novel method so that authenticity of individual blocks can be verified without sacrificing localization accuracy. The rest of the paper is organized as follows: in Sect. 2, we discuss the proposed countermeasures against the Holliman–Memon attack. The proposed watermarking method is described in Sect. 3. We present our simulation results and discussion in Sect. 4 and the conclusions are given in Sect. 5.

## 2 Countermeasures against the Holliman–Memon attack

In this section, we describe various countermeasure methods which have been proposed in the literature to resist the Holliman–Memon attack.

- Neighborhood dependent blocks

In [8], a practical approach is suggested to remove the block-wise independence of the watermark. The signature embedded in each block is calculated using some of the surrounded data from neighboring blocks, as well as the data within the block itself. Using this method, a collage of watermarked blocks cannot be authenticated by the detector, because the neighborhood relationship between the blocks is not preserved in the fake image. However, this introduces some ambiguity to the localization, because a change in one block can change the signature that should be embedded in its neighbors.

- Block index and image index in signature computation

Wong and Memon [9] suggested including a block index and a unique image index while computing the signature for each block. The use of block index solves the problem of block swapping in a watermarked image. The attacker needed to search for the most similar blocks only at identical block positions of all database images. Although this complicates the attacker's

task, it is still possible to launch an attack if the number of database images is high. The use of a unique image index completely eliminates the possibility of this attack. However, the image index is also necessary for verification at the detector.

In some applications it may be possible to make such indices publicly available; however, in many cases managing such indices would create additional overheads and may not be operational feasible. To solve this problem, the authors suggested extracting the image index from the image itself by computing the hash of its MSBs. However if any of the MSBs is altered due to tampering in the image, it will lead to a complete loss of localization. Another approach to solve the problem of image index management has been proposed by Fridrich et al. [10]. In this method, an image index is embedded in the image at multiple locations in a robust manner. In case of a tampering, the multiple copies increase the chance of extracting the correct image index at the detector. However, this does not guarantee the correct index extraction in all cases of tampering and the embedding process increases visual distortion in the watermarked image.

- Content origin and authentication

In [7], the information about the content origin is embedded in each block of the image. This method is based on Wong's scheme and the symmetric logo is embedded instead of any fixed logo. The symmetric logo contains information about image dimensions, camera ID, block index, author ID, image index and other ancillary data. The logo structure is used to verify the block integrity and the logo content provides the information about the block origin. Each block in the fake image after the Holliman–Memon attack is authenticated and the origin of such authentic blocks could be detected from the extracted logos.

- Hierarchical watermarking

Celik et al. [11] proposed a hierarchical watermarking method based on Wong's scheme. Using this method, the image is divided into blocks in a multilevel hierarchy and block signatures are calculated in each hierarchy. While signatures of small blocks on the lowest level of the hierarchy ensure superior tamper localization accuracy, higher level block signatures can resist this attack through a trade-off between security and the localization accuracy.

The main objective of the localization method is to verify authenticity of each block in an image. Wong's scheme achieves this objective against any kind of tampering with the accuracy of a chosen block size; however its drawback lies in verifying each block in the fake image to be authentic against the Holliman–Memon attack. The above discussion shows

that the proposed countermeasures could resist this attack with the performance reduction as compared to Wong's scheme. The countermeasure using the unique image index in [9] is of particular interest to this paper. If the correct image index can be extracted for verification after the fragile embedding process, then the Holliman–Memon attack could be resolved without loss of localization accuracy. In the next section, we address this motivation by proposing a novel localization method using a unique image index scheme. In the proposed method, the localization accuracy remains at the level of chosen block size and it is possible to determine the authenticity of each block in the fake image resulting from this attack.

### 3 Proposed method

In this section, we propose a new method for enhancing the security of Wong's scheme to resist the Holliman–Memon attack. The idea behind the method is based on the use of a unique image index in the computation of the authentication signature for every block. The image index is embedded in a fragile manner so that the visual distortion is minimized. By designing an informed detector, the correct image index is estimated from the fake image. The proposed embedding and detection methods are described as follows:

#### *Embedding*

1. The original image  $\mathbf{X}$  of size  $M \times N$  is partitioned into non-overlapping blocks of  $12 \times 12$  pixels. Let  $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_B\}$  denote the individual blocks in a sequential order, starting from left to right and top to bottom of the image, where  $B = (M \times N)/144$ . Watermarking is performed for each block independently and in a sequential order.
2. For each block  $\mathbf{X}_r$ , a corresponding block  $\mathbf{X}_r^e$  is formed by setting the least significant bit (LSB) of each pixel to zero. The 128-bit hashed message authentication code ( $\mathbf{H}_r$ ) is computed using the function  $\text{HMAC}(\cdot)$  and

$$\mathbf{H}_r = \text{HMAC}(\mathbf{X}_r^e, K, r, I_X) \quad (1)$$

where  $K, r, I_X$  denote secret key, block index and image index, respectively.

3. Out of the 144 least significant bits in a block,  $\mathbf{H}_r$  is inserted in 128 LSB positions and the rest 16 LSBs hold the image index. Using the 16-bit image index, it is possible to securely watermark  $2^{16}$  or 65536 images with one secret key.

*Detection* After embedding the authentication signature, the following side information of the watermarked image is available to the detector.

- Each block should be authenticated using the embedded image index.
- All blocks contain the same image index.
- All blocks are connected with each other, i.e. it is possible to move from one pixel to any other pixel in the image using an eight-connected path.

It is possible to estimate the correct image index at the detector by using the above side information.

1. The test image  $X'$  is partitioned into non-overlapping blocks of  $12 \times 12$  pixels and detection is performed for each block in a sequential order.
2. The 128-bit authentication signature  $H_r^d$  and the 16-bit image index  $I_X^d$  are extracted from the least significant bits of each block  $X_r^d$ .  $X_r^d$  is formed by setting LSBs of  $X_r'$  to zero. The HMAC ( $H_r^c$ ) is computed according to the following equation:

$$H_r^c = \text{HMAC}(X_r^d, K, r, I_X^d) \tag{2}$$

3. All the bits in a block can be divided into three groups; (a) MSBs (b) LSBs containing the authentication signature, and (c) LSBs containing the image index. If any of these bits is changed after an attack, either the extracted signature or the computed signature will be altered. A matrix ( $\mathbf{R}$ ) of  $(M/12)$  rows and  $(N/12)$  columns is constructed while computing  $H_r^c$  for each block. Each entry of  $\mathbf{R}$  represents a particular block in image  $X'$  and this relationship is shown in Fig. 1. The magnitude of an entry in  $\mathbf{R}$  is '1' if  $H_r^d$  and  $H_r^c$  matches exactly in the corresponding block; otherwise it is equal to '0'.
4. Different image indices can be extracted from all the blocks in  $X'$ . Let each such index be termed as the candidate image index ( $I_X^c$ ). For every  $I_X^c$ , the authenticity score ( $A_S$ ) is computed as follows:

**Algorithm**

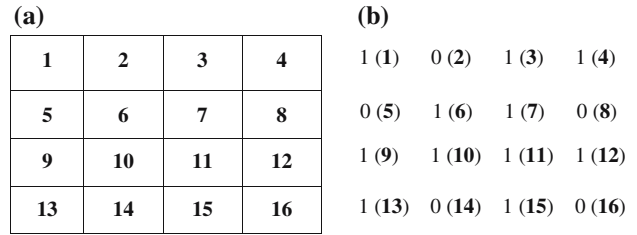
A matrix ' $\mathbf{T}$ ' of  $(M/12)$  rows and  $(N/12)$  columns is constructed and each entry of  $\mathbf{T}$  corresponds to an individual block in a sequential order like in  $\mathbf{R}$ .

for  $u = 1, 2, \dots, (M/12)$  and  $v = 1, 2, \dots, (N/12)$

if ( $\mathbf{R}(u, v) = 1$  and  $I_X^d = I_X^c$ ), then

$$\mathbf{T}(u, v) = 1$$

else



**Fig. 1** a An image of  $48 \times 48$  pixels partitioned into blocks of  $12 \times 12$  pixels and the block numbers are shown in a sequential order, b the matrix ' $\mathbf{R}$ ' having the entries either 0 or 1 and the corresponding block number are written within the bracket

$$\mathbf{T}(u, v) = 0$$

end

end

A score matrix ' $\mathbf{S}$ ' of  $(M/12)$  rows and  $(N/12)$  columns is then computed from  $\mathbf{T}$ .

for  $u = 1, 2, \dots, (M/12)$  and  $v = 1, 2, \dots, (N/12)$

if  $\mathbf{T}(u, v) = 1$ , then

$$\mathbf{S}(u, v) = (G + 1)/B \tag{3}$$

else

$$\mathbf{S}(u, v) = 0$$

end

end

The authenticity score for the candidate image index is,

$$A_S = \frac{1}{B} \sum_{u=1}^{(M/12)} \sum_{v=1}^{(N/12)} \mathbf{S}(u, v) \tag{4}$$

where  $G$  is the total number of 1's in  $\mathbf{T}$  that is eight-connected with the present entry at  $(u, v)$  and  $B$  is the total number of blocks.  $G$  is computed by treating  $\mathbf{T}$  as a binary image and then applying the connected component labeling procedure [12]. The matching between  $I_X^d$  and  $I_X^c$  is done on a bit-by-bit basis while computing  $\mathbf{T}$ .

5. The candidate image index with the highest authenticity score is chosen to be the estimated image index. Let the estimated image index be denoted as  $I_E$ . A block in the test image will be authenticated if the following conditions are satisfied:

- The image index extracted from the block ( $I_X^d$ ) is exactly equal to the estimated image index ( $I_E$ ).
- The corresponding entry in  $\mathbf{R}$  for the block is 1.

6. The authenticity measure ( $A_M$ ) of the test image is defined to quantify the attack severity and its value is equal to the authenticity score of  $I_E$ . The maximum

value of  $A_M$  is ‘1’ when all blocks in the test image are authenticated and ‘0’ when all blocks in the test image are inauthentic.

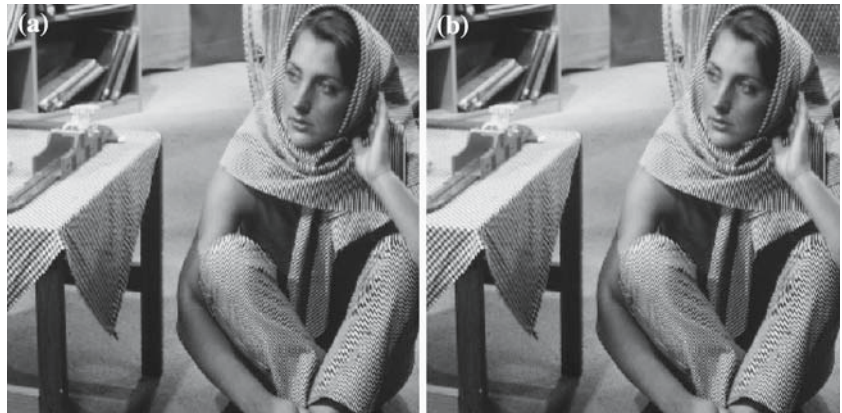
#### 4 Results and discussion

In this section, we present simulation results to demonstrate the effectiveness of the proposed method. In our implementation, we have used the HMAC as the authentication signature. The MD5 hash function [13] is used to compute the 128-bit HMAC. A digital signature can be embedded instead of the HMAC to allow public key verification of the watermarked image. In the first test case, we demonstrate the localization capability of this method against tampering in the watermarked image. The ‘Barbara’ image of size  $300 \times 300$  pixels is used as the original image. The decimal equivalent of the image index is 23159. Using the proposed method, the 128-bit authentication signature and 16-bit image index are embedded in 625 blocks of the original image. The original image and watermarked image are shown in Fig. 2. Without any tampering, all blocks in the

watermarked image are authenticated by the proposed detection method. The estimated image index is 23,159 and authenticity measure of the watermarked image is 1. The watermarked image is then tampered with the words ‘copyright image’ inserted in it. The resulting attacked image and authenticated image are shown in Fig. 3. The dark region in the authenticated image indicates the tampered blocks. The estimated image index is 23,159 and the authenticity measure of the attacked image is approximately 0.88.

The effectiveness of the proposed method against the Holliman–Memon attack is demonstrated in our second test case. As in [8], a database of fingerprint images is used for this attack. A total of 64 fingerprint images of size  $300 \times 300$  pixels are watermarked using the proposed embedding method. The images are watermarked using the 16-bit image indices whose decimal equivalent ranges from 1 to 64. The unwatermarked fingerprint image and the fake fingerprint image constructed using the Holliman–Memon attack is shown in Fig. 4. In the attack, the most similar block is searched at the identical block position of 64 database images using the mean square error (MSE) criterion. For the fake image,

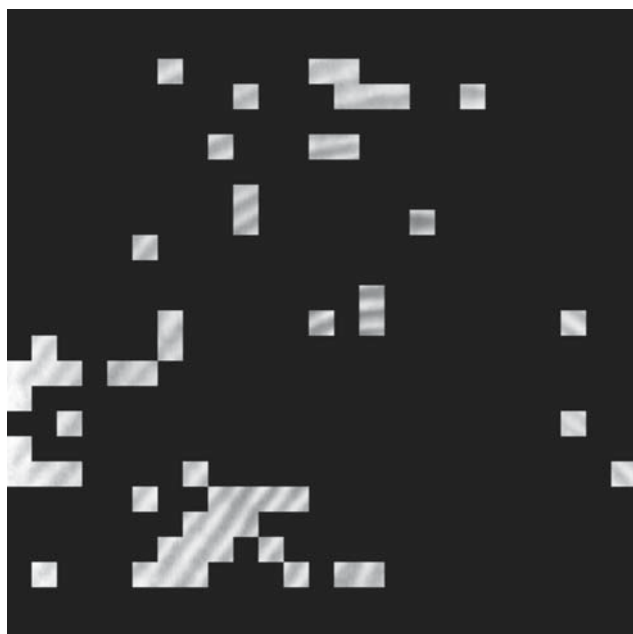
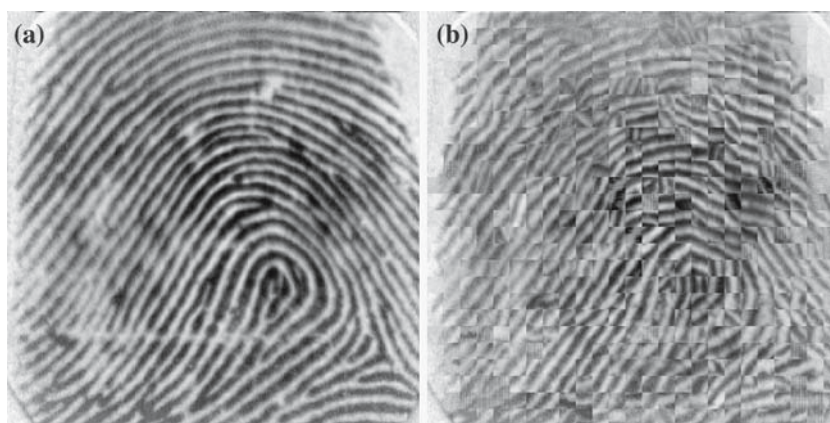
**Fig. 2** **a** Original ‘Barbara’ image of size  $300 \times 300$  pixels, **b** watermarked image after embedding the authentication signature in each block



**Fig. 3** **a** Attacked image in which the words ‘Copyright Image’ is inserted, **b** image showing tamper localization in dark regions

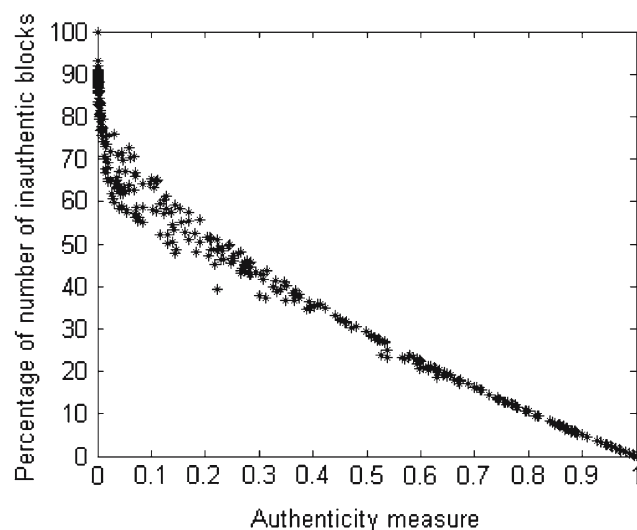


**Fig. 4** **a** Unwatermarked fingerprint image of size  $300 \times 300$  pixels, **b** fake image constructed using the Holliman–Memon attack



**Fig. 5** Detection output after verifying authenticity of the fake image using the proposed method. *Dark region* in the image shows the inauthentic blocks

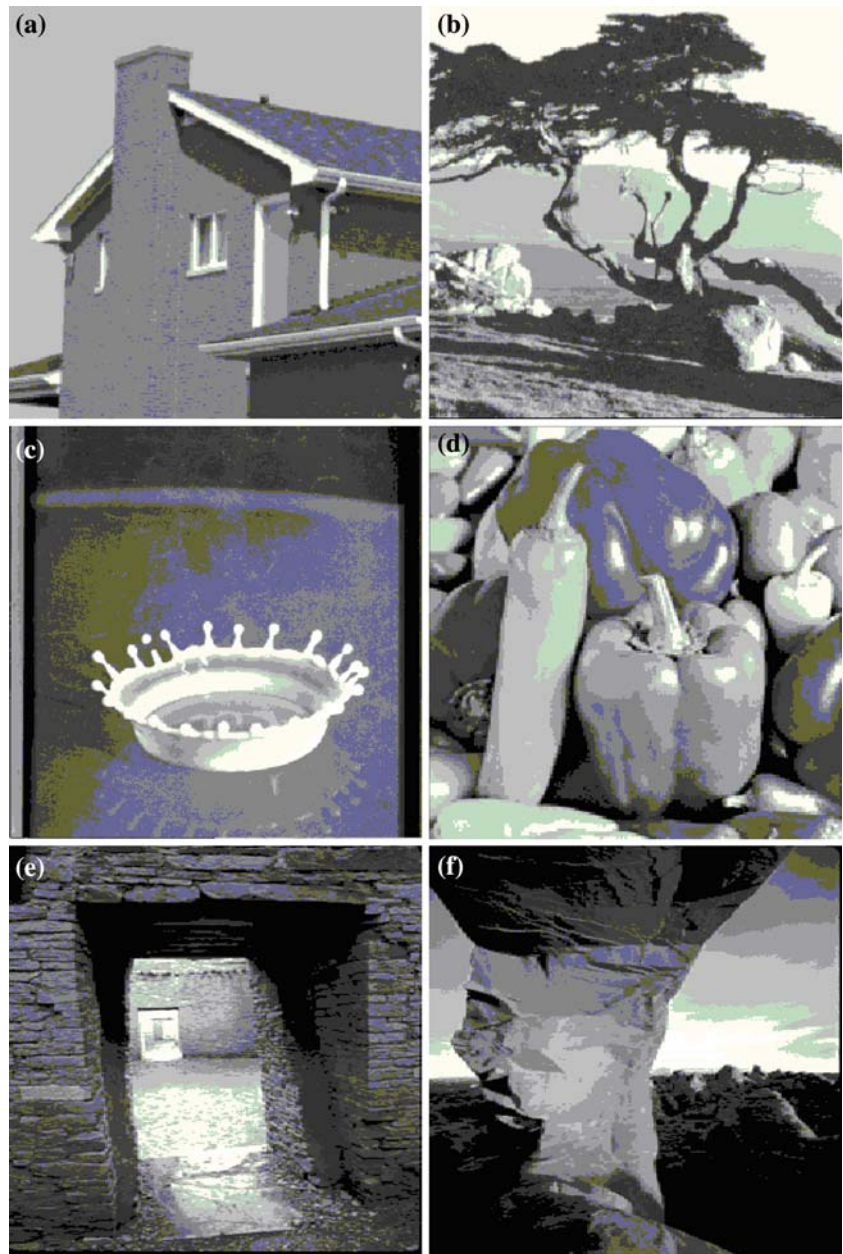
the PSNR is approximately 22.58 dB. Generally, two images are found to be perceptually similar when PSNR is approximately 40 dB. The low value of PSNR for the fake image indicates that the visual quality is degraded as compared to the unwatermarked fingerprint image. Several blocking artifacts are visible in the fake image. Since the block index is used in the signature computation step, the codebook construction domain is restricted in the proposed method as compared to the method in [8]. The visual quality of the fake image is therefore reduced. As the size of the database increases, it is possible to improve the visual quality of the fake image and the number of blocking artifacts may diminish. The proposed detection method is used to verify each block in the fake image and the result is shown in Fig. 5. A total of 55 blocks out of 625 blocks in the fake image are



**Fig. 6** Percentage of number of inauthentic blocks ( $P_I$ ) versus the authenticity measure ( $A_M$ ) for the fingerprint test image

authenticated. The estimated image index is 8 and the authenticity measure of the fake image is approximately  $0.94 \times 10^{-3}$  which indicates the attack severity.

To demonstrate the correlation between the proposed authenticity measure and attack severity, we perform the following experiment using 64 watermarked fingerprint images. Various fake images are generated for the fingerprint test image in Fig. 4 using the Holliman–Memon attack. For generating a fake image, a set of fingerprint images are randomly chosen from 64 images by using a key. A total of 381 fake images are generated by varying the key and the number of watermarked images used for performing the Holliman–Memon attack. The proposed method is used to verify the fake images and the relationship between percentage of number of inauthentic blocks ( $P_I$ ) and the authenticity measure ( $A_M$ ) is shown in Fig. 6. As  $P_I$  increases (attack severity increases), the authenticity measure has a decreasing trend and vice

**Fig. 7** Original test images

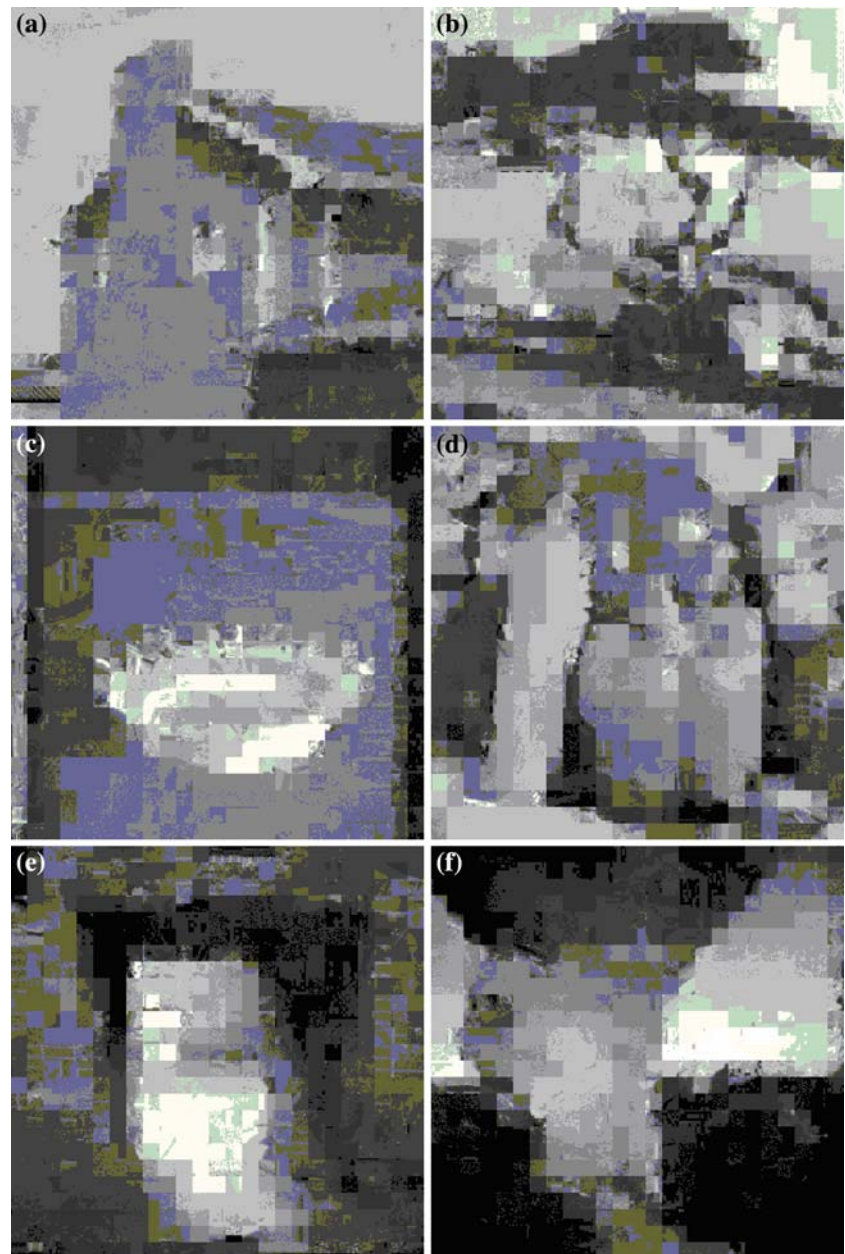
versa. The correlation coefficient [14] between  $P_I$  and  $A_M$  is found out to be  $-0.96$  approximately.

To test the effectiveness of the proposed method further, a total of 164 images are chosen from the databases [15,16] and the images are resized to generate the original images of size  $300 \times 300$  pixels. The proposed embedding method is used to generate 164 watermarked images using the 16-bit image indices whose decimal equivalent ranges from 1 to 164. Six original images are chosen for demonstrating the effectiveness of the proposed method and the images are shown in Fig. 7a–f. For each original image, the corresponding fake image is constructed using the Holliman–Memon attack. While constructing a fake image, the watermarked image

corresponding to the original image is not considered during the attack and thus each fake image is generated from 163 watermarked images. The fake images are shown in Fig. 8a–f. The proposed detection method is used to verify the fake images and the detection output is shown in Fig. 9a–f. The authenticity measure, PSNR and number of inauthentic blocks for the fake images are shown in Table 1. The detection of large portions of inauthentic regions and low authenticity measure for the fake images shows that the Holliman–Memon attack is practically infeasible against the proposed method.

To show the correlation between the authenticity measure and attack severity, various fake images are generated for each original image through the procedure

**Fig. 8** The fake images constructed by performing the Holliman–Memon attack using 163 watermarked images

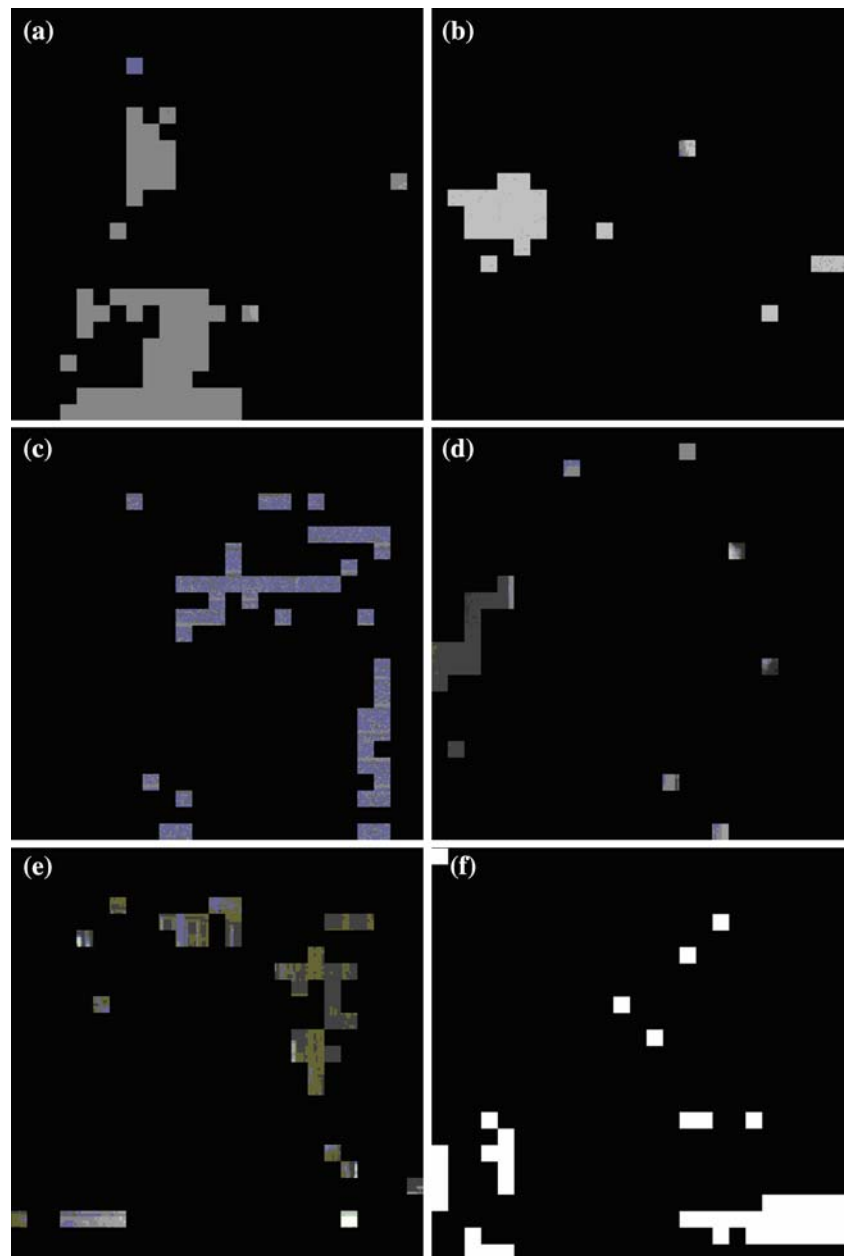


used in case of the fingerprint test image. Then the proposed detection method is used to verify the fake images. The relationship between percentage of number of inauthentic blocks and the authenticity measure is shown in Fig. 10a–f for six test cases. The high magnitude of the correlation coefficients between  $P_I$  and  $A_M$  summarized in Table 2 shows that the proposed authenticity measure quantifies the attack severity.

The performance of the proposed method to resist the Holliman–Memon attack can be compared with previous localization methods. As discussed in Sect. 2, the possibility of this attack is entirely eliminated using a unique image index in [9]. However, for verification

it is necessary to manage the database of such indices which may not be possible in many practical applications. In [11], the hierarchical watermarking method could detect this attack at a larger region instead of the chosen block size. In the proposed method, each block in the fake image can be verified without any loss or ambiguity in localization. After both Holliman–Memon attack and content tampering, localization accuracy of the proposed method remains at the level of chosen block size similar to Wong’s scheme. Since the correct image index can be estimated at the detector, it is not necessary to manage the database of image indices.

**Fig. 9** Detection output after verifying authenticity of the fake images. *Dark region* in the image shows the inauthentic blocks

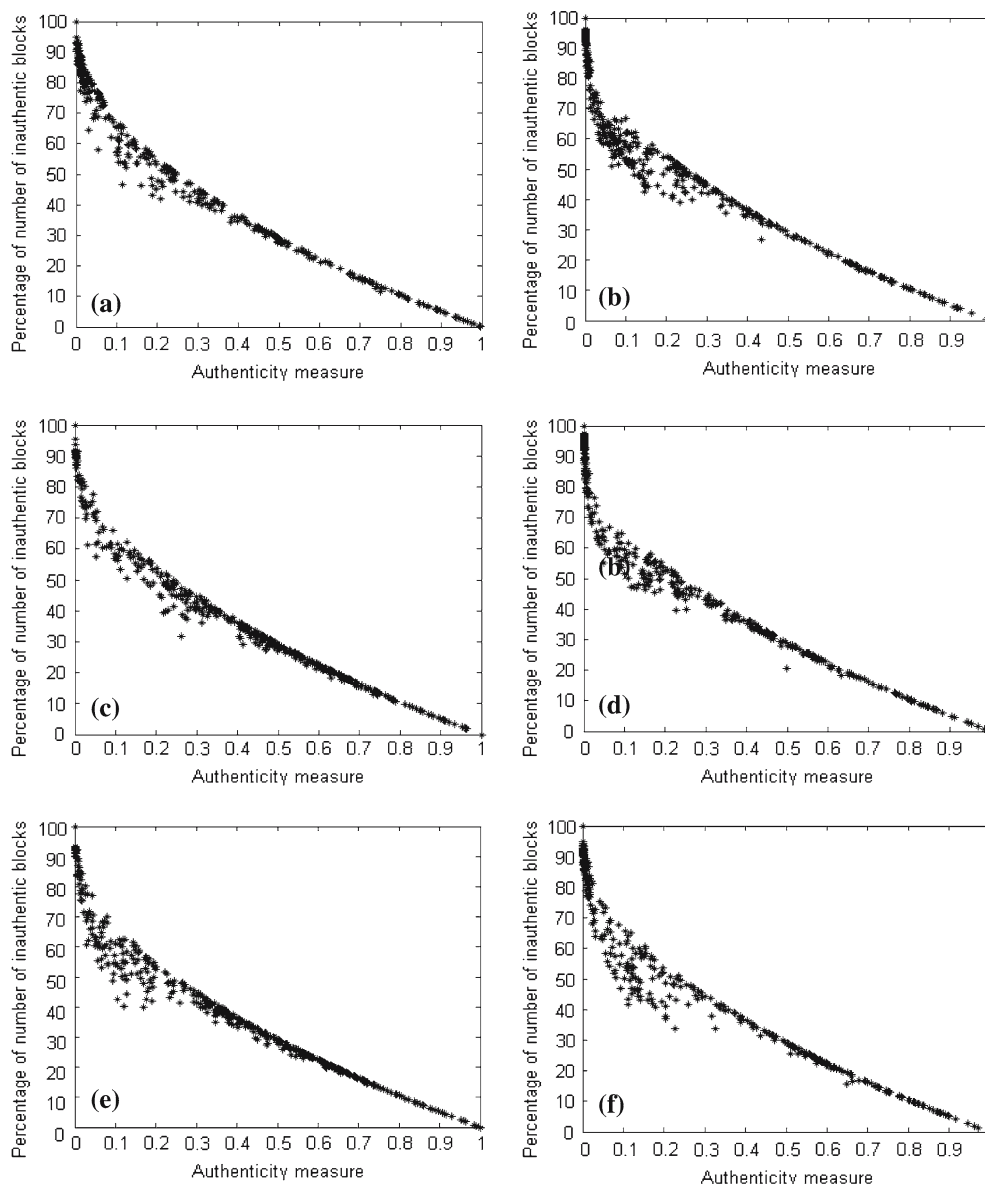


**Table 1** Performance attributes for all test cases

Test cases	PSNR (dB)	Number of inauthentic blocks	Authenticity measure
a	22.28	556	0.0069
b	18.21	600	$9.44 \times 10^{-4}$
c	21.55	575	0.0015
d	20.59	605	$4.5 \times 10^{-4}$
e	20.15	583	0.0011
f	21.15	581	0.0014

The authenticity measure quantifies the attack severity in an image by taking connectivity among the authentic blocks into account. As the attacker tries to

approximate an unwatermarked image by finding similar blocks in large number of database images, the authenticity measure for the fake image decreases significantly. For example, the authenticity measure for the fingerprint fake image in Fig. 4 is estimated to be  $0.94 \times 10^{-3}$  as compared to 1 as in the case of no attack. More blocks in the fake image are verified to be authentic when this measure is of high value. The blocks for the fake image are to be chosen within a less number of database images and the blocks from any such image should be connected with each other to maximize this measure. This would make the attacker's task more difficult to generate the fake image of reasonable perceptual quality. All blocks



**Fig. 10** Percentage of number of inauthentic blocks ( $P_I$ ) versus the authenticity measure ( $A_M$ ) for six test cases

**Table 2** Correlation coefficients between  $P_I$  and  $A_M$

Test cases	Number of fake images	Correlation coefficient
a	445	-0.96
b	449	-0.91
c	436	-0.95
d	417	-0.93
e	478	-0.94
f	414	-0.93

can be authenticated only if the authenticity measure of the fake image is 1. In that case, the fake image should be exactly equal to one of the database images.

Localization accuracy of the proposed method is bounded by the chosen block size. As the authentication signature is embedded into the least significant bit-plane of the block, minimum block size is determined by the length of the authentication signature. High security against content modification (the probability of undetected modification is  $2^{-128}$ ) is obtained by using the cryptographic hash function in the signature computation step. The only method to break the hash function is a brute-force attack which, according to the birthday paradox requires roughly  $2^{(n/2)}$  attempts to be successful, where  $n$  is the number of bits in the hash output.

In Wong’s scheme, a block size of  $8 \times 8$  pixels is chosen to embed the 64-bit signature. In this case approxi-

**Fig. 11** **a** Original image, **b** watermarked image



mately  $2^{32}$  attempts are needed to find a block whose hash output is same as the original block. This scale of computation is potentially feasible using present day's technology. For this reason, the HMAC of length 128 bits is used in the proposed method to attain cryptographic security. This implies that the smallest block size would be about  $12 \times 12$  pixels; thus the localization accuracy is equal to this block size. In [9], the chosen block size is also  $12 \times 12$  pixels for the 128-bit signature and the block size in [11] is  $10 \times 10$  pixels for the 64-bit signature. For this method, to accommodate the 128-bit signature and the payload of higher hierarchies, the block size would be greater than  $12 \times 12$  pixels. In cropping attack, a smaller rectangular region is selected within the whole image and the remaining portions discarded. Cropping is one of the image manipulations in which the watermark detection method fails to authenticate regions due to the loss of synchronization of block boundaries. To detect cropping in block-wise independent watermarking scheme, a sliding-window search is utilized to regain synchronization with the block boundaries [11]. This search method can be used in the proposed method to find block boundaries in the cropped image. The computational complexity of a localization method depends on the number of signature operations which is equal to total number of blocks in the image. Thus the proposed method has the same computational complexity as the previous schemes. In our implementation, both embedding and detection processes are performed approximately within 55 s for the original image of size  $300 \times 300$  pixels.

To show the performance of the proposed method against face tampering, we perform the following experiment. The original image [16] and watermarked image of size  $300 \times 300$  pixels is shown in Fig. 11. The watermarked image is tampered by replacing the face from another image [16] shown in Fig. 12. The tampered image and the detection output are shown in Fig. 13. Dark regions correctly indicate the tampered face of



**Fig. 12** Image used for face tampering

the image. The authenticity measure of the tampered image is approximately 0.6 and a total of 140 blocks out of 625 blocks in the tampered image are verified to be inauthentic. When the face of a person is changed, it largely changes the semantic meaning of the image. Thus the authenticity measure should be low after detection of face tampering. Using the proposed method, the authenticity measure for the tampered image is reduced from 1 to 0.6. To enhance the performance of the proposed method against face tampering, it is necessary to consider the semantic meaning of content while designing the authenticity measure. For that purpose, the authenticity of extracted watermarks from different regions in an image should be given unequal importance based on the semantic meaning of content. Such kind of authenticity measure computation for a fake

**Fig. 13** **a** Tampered image after replacing the face, **b** dark regions correctly indicate the tampered face in the detection output

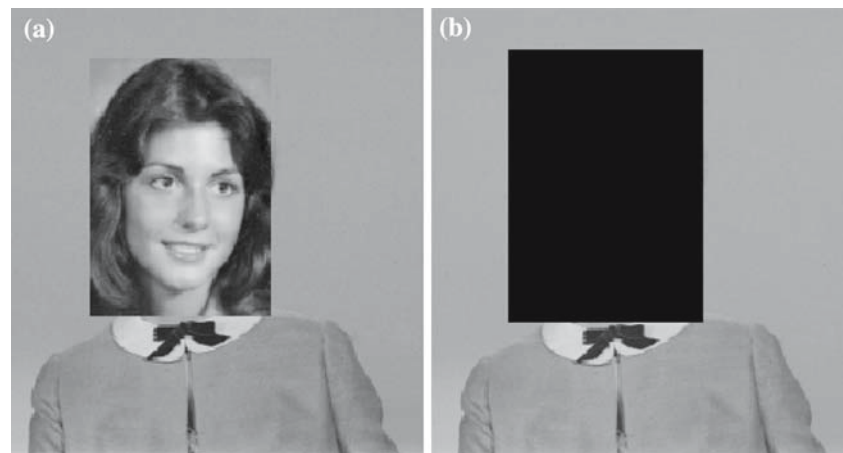


image without using the watermarked image constitutes a very interesting line for future research in content authentication and forensics. Another limitation of the proposed method is that it is particularly effective for fragile authentication purpose and is not robust against noise and unintentional signal processing operations.

## 5 Conclusion

In this paper, we proposed a new method in authentication watermarking to resist the Holliman–Memon attack. In this method, the authentication signature and a unique image index was embedded in the least significant bit-plane of the watermarked image to minimize visual distortion. By using side information about the watermarked image, the informed detector estimated the correct image index from the fake image. The image index estimation from the fake image can definitely be an alternative to keeping a directory of image indices. Authenticity of each block in the fake image was verified without any loss or ambiguity in localization. The localization accuracy of the proposed method remained at the chosen block size as compared to Wong's scheme. The authenticity measure was defined to quantify the attack severity and simulation results showed that the percentage of number of authentic blocks in a fake image was highly correlated to its authenticity measure. The attacker's task to generate the fake image became increasingly difficult with the inclusion of the authenticity measure. The proposed method is particularly useful in applications such as e-commerce, legal applications, medical archiving, news reporting and image database applications to address the image authentication issue. In these applications, it is possible to watermark a large number of images using a single key and many image indices, without any need for managing the indices. In terms of future work, we will focus on providing a for-

mal proof about the security of the proposed method against the Holliman–Memon attack.

## References

1. Swanson, M.D., Kobayashi, M., Tewfik, A.H.: Multimedia data-embedding and watermarking technologies. *Proc. IEEE* **86**(6), 1064–1087 (1998)
2. Cox, I.J., Miller, M.L.: A review of watermarking and the importance of perceptual modeling. *Proc. SPIE* **3016**, 92–99 (1999)
3. Menezes, A., van Oorschot, P., Vanstone, S.: *Handbook of Applied Cryptography*. CRC, Boca Raton, FL (1997)
4. Wong, P.: A Watermark for image integrity and ownership verification. In: *Proceedings at the IS&T PIC*, Portland, OR (1998)
5. Wong, P.: A public key watermark for image verification and authentication. In: *Proceedings at the IEEE International Conference Image Processing*, Chicago, pp. 425–429 (1998)
6. Yeung, M., Mintzer, F.: An invisible watermarking technique for image verification. In: *Proc. IEEE International Conference Image Processing*, Santa Barbara, pp. 680–683 (1997)
7. Fridrich, J.: Security of fragile authentication watermarks with localization. *Proc. SPIE* **4675**(75), 691–700 (2002)
8. Holliman, M., Memon, N.: Counterfeiting attacks on oblivious block-wise independent invisible watermarking schemes. *IEEE Trans. Image Process.* **9**(3), 432–441 (2000)
9. Wong, P.W., Memon, N.: Secret and public key image watermarking schemes for image authentication and ownership verification. *IEEE Trans. Image Process.* **10**(10), 1593–1601 (2001)
10. Fridrich, J., Goljan, M., Baldoza, A.C.: New fragile authentication watermark for images. In: *Proceedings of the IEEE International Conference on Image Processing*, Vancouver, pp. 446–449 (2000)
11. Celik, M.U., Sharma, G., Saber, E., Tekalp, A.M.: Hierarchical watermarking for secure image authentication with localization. *IEEE Trans. Image Process.* **11**(6), 585–595 (2002)
12. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 2nd Edn. Prentice Hall, Englewood Cliffs (2002)
13. Rivest, R.L.: RFC 1321: the MD5 Message-Digest Algorithm. Internet Activities Board (1992)
14. Anderson, D., Sweeney, D.J., Williams, T.A.: *Introduction to Statistics: An Applications Approach*. West Publishing Company, New York (1981)
15. Image database: [www.petitcolas.net/fabien/watermarking/image\\_database/index.html](http://www.petitcolas.net/fabien/watermarking/image_database/index.html)
16. The USC-SIPI image database: <http://sipi.usc.edu/database/index.html>